Attorney's Docket No.: 60-009132-US(PAR) PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Express Mail No.: EL067144491US
In re application of: LAURILA et al. **Group No.:**
Serial No.: 0 /
Filed: Herewith **Examiner:**
For: METHOD IN SPEECH RECOGNITION AND A SPEECH RECOGNITION DEVICE

Commissioner of Patents and Trademarks
Washington, D.C. 20231

TRANSMITTAL OF CERTIFIED COPY

Attached please find the certified copy of the foreign application from which priority is claimed for this case:

Country	**: Finland**
Application Number	**: 990078**
Filing Date	**: 18 January 1999**

WARNING: "When a document that is required by statute to be certified must be filed, a copy, including a photocopy or facsimile transmission of the certification is not acceptable." 37 CFR 1.4(f) (emphasis added.)

SIGNATURE OF ATTORNEY

Reg. No.: 24,622

Clarence A. Green

Type or print name of attorney

Tel. No.: (203) 259-1800

Perman & Green, LLP

P.O. Address

425 Post Road, Fairfield, CT 06430

NOTE: The claim to priority need be in no special form and may be made by the attorney or agent if the foreign application is referred to in the oath or declaration as required by § 1.63.

#2

CERTIFICATE

I, Tuulikki Tulivirta, hereby certify that, to the best of my knowledge and belief, the following is a true translation, for which I accept responsibility, of a certified copy of Finnish Patent Application 990078 filed on 18 January 1999.

Tampere, 16 December 1999

VIRALLINEN KÄÄNTÄJÄ
TUULIKKI TULIVIRTA
CERTIFIED TRANSLATOR

Tuulikki Tulivirta

Tuulikki Tulivirta
Certified Translator (Act 1148/88)

Tampereen Patenttitoimisto Oy
Hermiankatu 6
FIN-33720 TAMPERE
Finland

Method in speech recognition and a speech recognition device

The present method relates to a method in speech recognition as set forth in the preamble of the appended claim 1, a speech recognition device as set forth in the preamble of the appended claim 8, and a speech-controlled wireless communication device as set forth in the preamble of the appended claim 11.

For facilitating the use of wireless communication devices, speech recognition devices have been developed, whereby a user can utter speech commands which the speech recognition device attempts to recognize and convert to a function corresponding to the speech command, *e.g.* a command to select a telephone number. A problem in the implementation of speech control has been for example the fact that different users say the speech commands in different ways: the speech rate can be different between different users, so does the speech volume, voice tone, *etc.* Furthermore, speech recognition is disturbed by a possible background noise, whose interference outdoors and in a car can be significant. Background noise makes it difficult to recognize words and to distinguish between different words *e.g.* upon uttering a telephone number.

Some speech recognition devices apply a recognition method based on a fixed time window. Thus, the user has a predetermined time within which s/he must utter the desired command word. After the expiry of the time window, the speech recognition device attempts to find out which word/command was uttered by the user. However, such a method based on a fixed time window has *e.g.* the disadvantage that all the words to be uttered are not equally long; for example, in names, the given name is often clearly shorter than the family name. Thus, after a shorter word, more time will be consumed for the recognition than in the recognition of a longer word. This is inconvenient for the user. Furthermore, the time window must be set according to slower speakers so that recognition will not be started until the whole word is uttered. When words are uttered faster, a delay between the uttering and the recognition increases the inconvenient feeling.

Another known speech recognition method is based on patterns formed of speech signals and their comparison. Patterns formed of command words are stored beforehand, or the user may have taught desired words which have been formed into patterns and stored. The speech recognition device compares the stored patterns with feature vectors formed of sounds uttered by the user during the utterance and calculates the probability for the different words (command words) in the vocabulary of the speech recognition device. When the probability for a command word exceeds a predetermined value, the speech recognition device selects this command word as the recognition result. Thus, incorrect recognition results may occur particularly in the case of words in which the beginning resembles phonetically another word in the vocabulary. For example, the user has taught the speech recognition device the words "Mari" and "Marika". When the user is saying the word "Marika", the speech recognition device may make "Mari" as the recognition decision, even though the user may not yet have had time to articulate the end of the word. Such speech recognition devices typically use the so-called Hidden Markov Model (HMM) speech recognition method.

U.S. patent 4,870,686 presents a speech recognition method and a speech recognition device, in which the determination of the end of words by the user is based on silence; in other words, the speech recognition device examines if there is a perceivable audio signal or not. A problem in this solution is the fact that a too loud background noise may prevent the detection of pauses, wherein the speech recognition is not successful.

It is an aim of the present invention to provide an improved method for detecting pauses in speech and a speech recognition device. The invention is based on the idea that a tone band to be examined is divided into sub-bands, and the power of the signal is examined in each sub-band. If the power of the signal is below a certain limit in a sufficient number of sub-bands for a sufficiently long time, it is deduced that there is a pause in the speech. The method of the present invention is characterized in what will be presented in the characterizing part of the appended claim 1. The speech recognition device according to the present invention is characterized in what will be presented in the char-

acterizing part of the appended claim 8. The wireless communication device of the present invention is characterized in what will be presented in the characterizing part of the appended claim 11.

The present invention gives significant advantages to the solutions of prior art. By the method of the invention, a more reliable detection of a gap between words can be obtained than by methods of prior art. Thus, the reliability of the speech recognition is improved and the number of incorrect and failed recognitions is reduced. Furthermore, the speech recognition device is more flexible with respect to manners of speaking by different users, because the speech commands can be uttered more slowly or faster without an inconvenient delay in the recognition or recognition taking place before an utterance has been completed.

By the division into sub-bands according to the invention, it is possible to reduce the effect of external interference. Spurious signals *e.g.* in a car have typically a relatively low frequency. In solutions of prior art, the energy contained in the whole frequency range of the signal is utilized in the recognition, wherein signals which are strong but have a narrow band width reduce the signal-to-noise ratio to a significant degree. Instead, if the frequency range to be examined is divided into sub-bands according to the invention, the signal-to-noise ratio can be improved significantly in such sub-bands in which the proportion of spurious signals is relatively small, which improves the reliability of the recognition.

In the following, the present invention will be described in more detail with reference to the appended drawings, in which

Fig. 1 is a flow chart illustrating the method according to an advantageous embodiment of the invention,

Fig. 2 is a reduced flow chart showing the speech recognition device according to an advantageous embodiment of the invention,

Fig. 3 is a state machine chart illustrating rank-order filtering to be applied in the method according to an advantageous embodiment of the invention, and

Fig. 4 is a flow chart illustrating the logic for deducing a pause to be applied in the method according to an advantageous embodiment of the invention.

The following is a description on the function of the method according to an advantageous embodiment of the invention, with reference to the flow chart of Fig. 1 and using as an example a speech-controlled wireless communication device MS according to the flow chart of Fig. 2. In the speech recognition, an acoustic signal (speech) is converted, in a way known as such, into an electrical signal by a microphone, such as a microphone 1a in the wireless communication device MS or a microphone 1b in a hands-free facility 2. The frequency response of the speech signal is typically limited to the frequency range below 10 kHz, *e.g.* in the frequency range from 100 Hz to 10 kHz. However, the frequency response of speech is not constant in the whole frequency range, but there are more lower frequencies than higher frequencies. Furthermore, the frequency response of speech is different for different persons. In the method of the invention, the frequency range to be examined is divided into narrower sub-frequency ranges (M number of sub-bands). This is represented by block 101 in the appended Fig. 1. These sub-frequency ranges are not made equal in width but taking into account the characteristic features of the speech, wherein some of the sub-frequency ranges are narrower and some are wider. At the low frequencies characteristic of speech, the division is denser, *i.e.* the sub-frequency ranges are narrower than for the higher frequencies, which frequencies are more rare in speech. This idea is also applied in the Mel frequency scale, known as such, in which the width of frequency bands is based on the logarithmic function of frequency.

In connection with the division into sub-bands, the signals of the sub-bands are converted to a smaller sample frequency, *e.g.* by under-sampling or by low-pass filtering. Thus, samples are transferred from the block 101 to further processing at this lower sampling frequency. This sampling frequency is advantageously *ca.* 100 Hz, but it is obvious

that also other sampling frequencies can be applied within the scope of the present invention. These samples are converted into said feature vectors.

A signal formed in the microphone 1a, 1b is amplified in an amplifier 3a, 3b and converted into digital form in an analog-to-digital converter 4. The precision of the analog-to-digital conversion is typically in the range from 12 to 32 bits, and in the conversion of a speech signal, samples are taken advantageously 8'000 to 14'000 times a second, but the invention can also be applied at other sampling rates. In the wireless communication device MS of Fig. 2, the sampling is arranged to be controlled by a controller 5. The audio signal in digital form is transferred to a speech recognition device 16 which is in a functional connection with the wireless communication device 16 and in which different stages of the method according to the invention are processed. The transfer takes place *e.g.* via interface blocks 6a, 6b and an interface bus 7. In practical solutions the speech recognition device 16 can as well be arranged in the wireless communication device 16 itself or in another speech-controlled device, or as a separate auxiliary device or the like.

The division into sub-bands is made preferably in a first filter block 8, to which the signal converted into digital form is conveyed. This first filter block 8 consists of several band-pass filters which are in this advantageous embodiment implemented with digital technique and whose frequency ranges and band widths of the pass band differ from each other. Thus each band filtered part of the original signal passes the respective band-pass filter. For clarity, these band-pass filters are not shown separately in Fig. 2. These band-pass filters are implemented advantageously in the application software of a digital signal processor (DSP) 13, which is known as such.

At the next stage 102, the number of sub-bands is reduced preferably by decimating in a decimating block 9, wherein L number of sub-bands are formed ($L < M$), their energy levels being measurable. On the basis of the signal power levels of these sub-frequency ranges, it is possible to determine the signal energy in each sub-band. Also, the decimating

block 9 can be implemented in the application software of the digital signal processor 13.

An advantage obtained by the division into M sub-bands according to the block 1 is that the values of these M different sub-bands can be utilized in the recognition to verify the recognition result particularly in an application using coefficients according to the Mel frequency scale. However, the block 101 can also be implemented by forming directly L sub-bands, wherein the block 102 will not be necessary.

A second filter block 10 is provided for low pass filtering of signals of the sub-bands formed at the decimating stage (stage 103 in Fig. 1), wherein short changes in the signal strength are filtered off and they cannot have a significant effect in the determination of the energy level of the signal in further processing. After the filtration, a logarithmic function of the energy level of each sub-band is calculated in block 11 (stage 104) and the calculation results are stored for further processing in sub-band specific buffers formed in memory means 14 (not shown). These buffers are advantageously of the so-called FIFO type (First In - First Out), in which the calculation results are stored as figures of *e.g.* 8 or 16 bits. Each buffer accommodates N calculation results. The value N depends on the application in question. Thus, the calculation results p(t) stored in the buffer represent the filtered, logarithmic energy level of the sub-band at different measuring instants.

An arrangement block 12 performs so-called rank order filtering for the calculation results (stage 105), in which the mutual rank of the different calculation results are compared. At this stage 105, it is examined in the sub-bands whether there is possibly a pause in the speech. This examination is shown in a state machine chart in Fig. 3. The operations of this state machine are implemented substantially in the same way for each sub-band. The different functional states S0, S1, S2, S3 and S4 of the state machine are illustrated with circles. Inside these state circles are marked the operations to be performed in each functional state. The arrows 301, 302, 303, 304 and 305 illustrate the transitions from one functional state to another. In connection with these arrows are marked the criteria, whose realization will set off this transition. The curves 306, 307 and 308 illustrate the situation in which the functional

state is not changed. Also these curves are provided with the criteria for maintaining the functional state.

In the functional states S1, S2 and S3, a function f() is shown, which represents the performing of the following operations in said functional states: preferably N calculation results p(t) are stored in the buffer, and the lowest maximum value p_min(t) and the highest minimum value p_min(t) are determined advantageously by the following formulae:

$$p\_min(t) = min\left[max\left\langle p(i-N+1), p(i-N+2), \ldots, p(i)\right\rangle\right], \quad i = N, N+1, \ldots, t$$
$$p\_max(t) = max\left[min\left\langle p(i-N+1), p(i-N+2), \ldots, p(i)\right\rangle\right], \quad i = N, N+1, \ldots, t$$

Consequently, in the function f(t), the maximum value p_max(t) searched is the highest minimum value and the minimum value p_min(t) is the lowest maximum value of the calculation results p(i) stored in the different sub-band buffers. After this, the median power $p(t)_m$ is calculated, which is the median value of the calculation results p(t) stored in the buffer, and a threshold value thr by the formula $thr = p\_min + k \cdot (p\_max - p\_min)$, in which $0 < k < 1$. Next, in the function f(), a comparison is made between the median power $p(t)_m$ and the threshold value calculated above. The result of the calculation will set off different operations depending on the functional state in which the state machine is at a given time. This will be described in more detail hereinbelow in connection with the description of the different functional states.

After storing a group of sub-band specific calculation results p(t) of the speech (N results per sub-band), the speech recognition device will move on to execute said state machine, which is implemented in the application software of either the digital signal processor 13 or the controller 5. The timing can be made in a way known as such, preferably with an oscillator, such as a crystal oscillator (not shown). The executing is started from the state S0, in which the variables to be used in the state machine are set in their initial values (init()): a pause counter C is set to zero, the power minimum p_min at the starting moment t = 1 (p_min(t = 1)) is set to the theoretical value of ∞, in practice to the highest possible numerical value available in the speech recognition device.

This maximum value is influenced by the number of bits these power values are calculated with. Correspondingly, the power maximum p_max at the starting moment t = 1 (p_max (t = 1)) is set to the theoretical value of $-\infty$, in practice to the lowest possible numerical value available in the speech recognition device.

After setting of the initial values, the function moves on to the state S1, in which the operations of said function f() are performed, wherein *e.g.* the power minimum p_min and the power maximum p_max as well as the median power $p(t)_m$ are calculated. In the functional state S1, also the pause counter C is increased by one. This functional state prevails until the expiry of a predetermined initial delay. This is determined by comparing the pause counter C with a predetermined beginning value BEG. At the stage when the pause counter C has reached the beginning value BEG, the operation moves on to state S2.

In the functional state S2, the pause counter C is set to zero and the operations of the function f() are performed, such as storing of the new calculation result p(t), and calculation of the power minimum p_min, the power maximum p_max as well as the median power $p(t)_m$ and the threshold value thr. The calculated threshold value and the median power are compared with each other, and if the median power is smaller than the threshold value, the operation moves on to state S3; in other cases, the functional state is not changed but the above-presented operations of this functional state S2 are performed again.

In the functional state S3, the pause counter C is increased by one and the function f() is performed. If the calculation indicates that the median power is still smaller than the threshold value, the value of the pause counter C is examined to find out if the median power has been below the power threshold value for a certain time. Expiry of this time limit can be found out by comparing the value of the pause counter C with an utterance time limit END. If the value of the counter is greater than or equal to said expiry time limit END, this means that no speech can be detected on said sub-band, wherein the state machine is exited.

However, if the comparison of the threshold value and the median power in the functional state S3 showed that the median power ex-

ceeded the power threshold value, it can be deduced that speech is detected on this sub-band, and the state machine returns to the functional state S2, in which *e.g.* the pause counter C is reset and the calculation is started from the beginning.

Consequently, the operation of a state machine to be used in the method according to an advantageous embodiment of the invention was described above in a general manner. In a speech recognition device according to the invention, the above-presented functional stages are performed separately for each sub-band.

Sampling a speech signal is performed advantageously at intervals, wherein the stages 101—104 are performed after the calculation of each feature vector, preferably at intervals of *ca.* 10 ms. Correspondingly, in the state machine of each sub-band, the operations according to the each active functional state are performed once (one calculation time), *e.g.* in state S3 the pause counter C(s) of the sub-band in question is increased, the function f(s) is performed, wherein *e.g.* a comparison is made between the median power and the threshold value, and on the basis of the same, the functional state is either retained or changed.

After one calculating round has been performed for the state machines of all the sub-bands, the operation moves on to stage 106 in the speech recognition, wherein it is examined on the basis of the information received from the different sub-bands whether a sufficiently long pause has been detected in the speech. This stage 106 is illustrated as a flow chart in the appended Fig. 4. For clarifying the examination, some comparison values are determined, which are given initial values preferably in connection with the manufacture of the speech recognition device, but if necessary, these initial values can be changed according to the application in question and the usage conditions. The setting of these initial values is illustrated with block 401 in the flow chart of Fig. 4:

- activity threshold SB_ACTIVE_TH whose value is greater than zero but smaller than the detection time limit END,
- detection quantity SB_SUFF_TH whose value is greater than zero but smaller than or equal to the number L of sub-bands,

— minimum number SB_MIN_TH of sub-bands whose value is greater than zero but smaller than the detection quantity SB_SUFF_TH.

In the method according to the invention, to detect a pause in speech it is examined, on how many sub-bands the energy level has possibly remained below said power threshold value and for how long. As disclosed in the functional description of the state machine above, the pause counter C indicates how long the audio energy level has remained below the power threshold value. Thus, the value of the counter is examined for each sub-band. If the value of the counter is greater than or equal to the detection time limit END (block 402), this means that the energy level of the sub-band has remained below the power threshold value so long that a decision on detecting a pause can be made for this sub-band, *i.e.* a sub-band specific detection is made. Thus, the detection counter SB_DET_NO is preferably increased by one.

If the value of the counter is greater than or equal to the activity threshold SB_ACTIVE_TH (block 404), the energy level on this sub-band has been below the power threshold value thr for a moment but not yet a time corresponding to the detection time limit END. Thus, the activity counter SB_ACT_NO in block 405 is increased preferably by one. In other cases, there is either an audio signal on the sub-band, or the level of the audio signal has been below the power threshold value thr for only a short time.

Next, the operation moves on to block 406, in which the sub-band counter i used as an auxiliary variable is increased by one. On the basis of the value of this sub-band counter i, it can be deduced if all the sub-bands have been examined (block 407).

When the comparisons to the said pause counters have been made, it is examined, on how many sub-bands a pause was detected (the pause counter was greater than or equal to the detection time limit END). If the number of such sub-bands is greater than or equal to the detection quantity SB_SUFF_TH (block 408), it is deduced in the method that there is a pause in the speech (pause detection decision, block 409),

and it is possible to move on to the actual speech recognition to find out what the user uttered. However, if the number of sub-bands is smaller than the detection quantity SB_SUFF_TH, it is examined, if the number of sub-bands including a pause is greater than or equal to the minimum number of sub-bands SB_MIN_TH (block 410). Furthermore, it is examined in block 411 if any of the sub-bands is active (the pause counter was greater than or equal to the activity threshold SB_ACTIVE_TH but smaller than the detection time limit END). In the method according to the invention, a decision is made in this situation that there is a pause in the speech if none of the sub-bands is active.

In a noise situation, noise on some sub-bands may effect that a detection decision cannot be made on all sub-bands even though there were a pause in the speech that should be detected. Thus, by means of said sub-band minimum SB_MIN_TH, it is possible to verify the detection of a pause in the speech particularly under noise conditions. Thus, in a noise situation, if a pause is detected on at least said minimum number SB_MIN_TH of sub-bands, a pause is detected in the speech if the pause detection decision on these sub-bands remains in force for the duration of said detection time limit END.

Correspondingly, under good conditions, using said detection time limit END may prevent a too quick decision on detecting a pause. Under good conditions, the said minimum number of sub-bands can quickly cause a pause detection decision, even though there is no such pause in the speech to be detected. By waiting the detection time limit for substantially all of the sub-bands, it is verified that there is actually a pause in the speech.

In another advantageous embodiment of the invention, it is not examined before making the decision of detecting a pause whether any of the sub-bands is active. Thus, the decision on detecting a pause is made on the basis of the results of the comparisons presented above.

The operations presented above can be implemented advantageously *e.g.* in the application software of the controller or digital signal processor of the speech recognition device.

The above-presented method for detecting a pause in speech according to the advantageous embodiment of the invention can be applied at the stage of teaching a speech recognition device as well as at the stage of speech recognition. At the teaching stage, the disturbance conditions can be usually kept relatively constant. However, when a speech-controlled device is used, the quantity of background noise and other interference can vary to a great extent. For improving the reliability of speech recognition particularly under varying conditions, the method according to another advantageous embodiment of the invention is supplemented with adaptivity to the calculation of the threshold value thr. For achieving this adaptivity, a modification coefficient UPDATE_C is used, whose value is preferably greater than zero and smaller than one. The modification coefficient is first given an initial value within said value range. This modification coefficient is updated during speech recognition preferably in the following way. On the basis of the samples of the sub-bands stored in the buffers, a maximum power level win_max and a minimum power level win_min are calculated. After this, said calculated maximum power level win_max is compared with the power maximum p_max at the time, and said calculated minimum power level win_min is compared with the power minimum p_min. If the absolute value of the difference between the calculated maximum power level win_max and the power maximum p_max, or the absolute value of the difference between the calculated minimum power level win_min and the power minimum p_min has increased from the previous calculation time, the modification coefficient UPDATE_C is increased. On the other hand, if the absolute value of the difference between the calculated maximum power level win_max and the power maximum p_max, or the absolute value of the difference between the calculated minimum power level win_min and the power minimum p_min has decreased from the previous calculation time, the modification coefficient UPDATE_C is reduced. After this, a new power maximum and a new power minimum are calculated as follows:

$$p\_min(t) = (1 - UPDATE\_C) \cdot p\_min(t-1) + (UPDATE\_C \cdot win\_min)$$
$$p\_max(t) = (1 - UPDATE\_C) \cdot p\_max(t-1) + (UPDATE\_C \cdot win\_max)$$

The calculated new power maximum and minimum values are used at the next sampling round *e.g.* in connection with the performing of the function f(). The determination of this adaptive coefficient has *e.g.* the advantage that changes in the environmental conditions can be better taken into account in the speech recognition and the detection of a pause becomes more reliable.

The above-presented different operations for detecting a pause in the speech can be largely implemented in the application software of the controller and/or the digital signal processor of the speech recognition device. In the speech recognition device according to the invention, some of the functions, such as the division into sub-bands, can also be implemented with analog technique, which is known as such. In connection with the execution of the method, in the storing of the calculation results to be made at different stages, the variables, *etc.,* it is possible to use the memory means 14 of the speech recognition device, preferably a random access memory (RAM), a non-volatile random access memory (NVRAM), a FLASH memory, *etc.* The memory means 22 of the wireless communication device can as well be used for storing information.

Fig. 2, showing a the wireless communication device MS according to an advantageous embodiment of the invention, additionally shows a keypad 17, a display 18, a digital-to-analog converter 19, a headphone amplifier 20a, a headphone 21, a headphone amplifier 20b for a hands-free function 2, a headphone 21b, and a high-frequency block 23, all known *per se.*

The present invention can be applied in connection with several speech recognition systems functioning by different principles. The invention improves the reliability of detection of pauses in speech, which ensures the recognition reliability of the actual speech recognition. Using the method according to the invention, it is not necessary to perform the speech recognition in connection with a fixed time window, wherein the recognition delay is substantially not dependent on the rate at which the user utters speech commands. Also, the effect of background noise on speech recognition can be made smaller upon applying the method of the invention than is possible in speech recognition devices of prior art.

It is obvious that the invention is not limited solely to the embodiments presented above, but it can be modified within the scope of the appended claims.

Claims:

1. A method for detecting pauses in speech in speech recognition, in which method, for recognizing speech commands uttered by the user, the voice is converted into an electrical signal, **characterized** in that in the method, the frequency spectrum of the electrical signal is divided into two or more sub-bands, samples of the signals in the sub-bands are stored at intervals, the energy levels of the sub-bands are determined on the basis of the stored samples, a power threshold value (thr) is determined, and the energy levels of the sub-bands are compared with said power threshold value (thr), wherein the comparison results are used for producing a pause detecting result.

2. The method according to claim 1, **characterized** in that a detection time limit (END) and a detection quantity (SB_SUFF_TH) are determined, wherein in the method, the calculation of the length of a pause in a sub-band is started when the energy level of the sub-band falls below said power threshold value (thr), wherein in the method, a sub-band specific detection is performed when the calculation reaches the detection time limit (END), it is examined on how many sub-bands the energy level was below the power threshold value (thr) longer than the time detection limit (END), wherein a pause detection decision is made if the number of sub-band specific detections is greater than or equal to the detection quantity (SB_SUFF_TH).

3. The method according to claim 2, **characterized** in that in the method, also an activity time limit (SB_ACTIVE_TH) and an activity quantity (SB_MIN_TH) are determined, wherein a pause detection decision is made if the quantity of sub-band specific detections is greater than or equal to the activity quantity (SB_MIN_TH) and the activity time limit (SB_ACTIVE_TH) has not been reached on the other sub-bands in the calculation of the length of the pause in the sub-band.

4. The method according to claim 1, 2 or 3, **characterized** in that the power threshold value (thr) is calculated by the formula

$thr = p\_min + k \cdot (p\_max - p\_min)$, in which

p_min = the smallest power maximum determined of the stored samples of the sub-bands, and

p_max = the greatest power minimum determined of the stored samples of the sub-bands.

5. The method according to any of the claims 1 to 4, **characterized** in that said power threshold value (thr) is calculated adaptively by taking into account the environmental noise level at each instant.

6. The method according to claim 5, **characterized** in that for calculating said power threshold value (thr), a modification coefficient (UPDATE_C) is determined, and on the basis of the stored samples, the greatest power level (win_max) and the smallest power level (win_min) of the sub-bands are calculated, wherein the power maximum (p_max) and power minimum (p_min) are determined by the formulae:

$$p\_max(i,t) = (1 - UPDATE\_C) \cdot p\_max(i,t-1) + (UPDATE\_C \cdot win\_max)$$
$$p\_min(i,t) = (1 - UPDATE\_C) \cdot p\_min(i,t-1) + (UPDATE\_C \cdot win\_min)$$

in which $0 < UPDATE\_C < 1$,
$0 < i < L$, and
L is the number of sub-bands.

7. The method according to claim 6, **characterized** in that further in the method,

— the modification coefficient (UPDATE_C) is increased, if the absolute value of the difference between said calculated highest power level (win_max) and the power maximum (p_max), or the absolute value of the difference between said calculated lowest power level (win_min) and the power minimum (p_min) has increased,

— the modification coefficient (UPDATE_C) is reduced, if the absolute value of the difference between said calculated highest power level (win_max) and the power maximum (p_max), or the absolute value of the difference between said calculated lowest power level (win_min) and the power minimum (p_min) has decreased.

8. A speech recognition device (16) comprising means (1a, 1b) for converting speech commands uttered by a user into an electrical signal, **characterized** in that it also comprises:

- means (8) for dividing the frequency spectrum of the electrical signal into two or more sub-bands,
- means (14) for storing samples of the signals of the sub-bands at intervals,
- means (5, 13) for determining energy levels of the sub-bands on the basis of the stored samples,
- means (5, 13) for determining a power threshold value (thr),
- means (5, 13) for comparing the energy levels of the sub-bands with said power threshold value (thr), and
- means (5, 13) for detecting a pause in the speech on the basis of said comparison results.

9. The speech recognition device (16) according to claim 8, **characterized** in that the power threshold value is calculated by the formula

$thr = p\_min + k \cdot (p\_max - p\_min)$, in which

p_min = the smallest determined power maximum of the stored samples of the sub-bands, and

p_max = the greatest determined power minimum of the stored samples of the sub-bands.

10. The speech recognition device (16) according to claim 8 or 9, **characterized** in that it comprises also means (10, 11) for filtering the signals of the sub-bands before storage.

11. A wireless communication device (MS) comprising means (16) for recognizing speech and means (1a, 1b) for converting speech commands uttered by a user into an electrical signal, **characterized** in that the means (16) for recognizing speech comprise also:

- means (8) for dividing the frequency spectrum of the electrical signal into two or more sub-bands,
- means (14) for storing samples of the signals of the sub-bands at intervals,

- means (5, 13) for determining energy levels of the sub-bands on the basis of the stored samples,
- means (5, 13) for determining a power threshold value (thr),
- means (5, 13) for comparing the energy levels of the sub-bands with said power threshold value (thr), and
- means (5, 13) for detecting a pause in the speech on the basis of said comparison results.

Abstract

In a method for detecting pauses in speech in speech recognition, for recognizing speech commands uttered by the user, the voice is converted into an electrical signal, whose frequency spectrum is divided into two or more sub-bands. Samples of the signals on the sub-bands are stored at intervals, the energy levels of the sub-bands are determined on the basis of the stored samples, a power threshold value (thr) is determined, and the energy levels of the sub-bands are compared with said power threshold value (thr). The comparison results are used for producing a pause detecting result.

Fig. 1

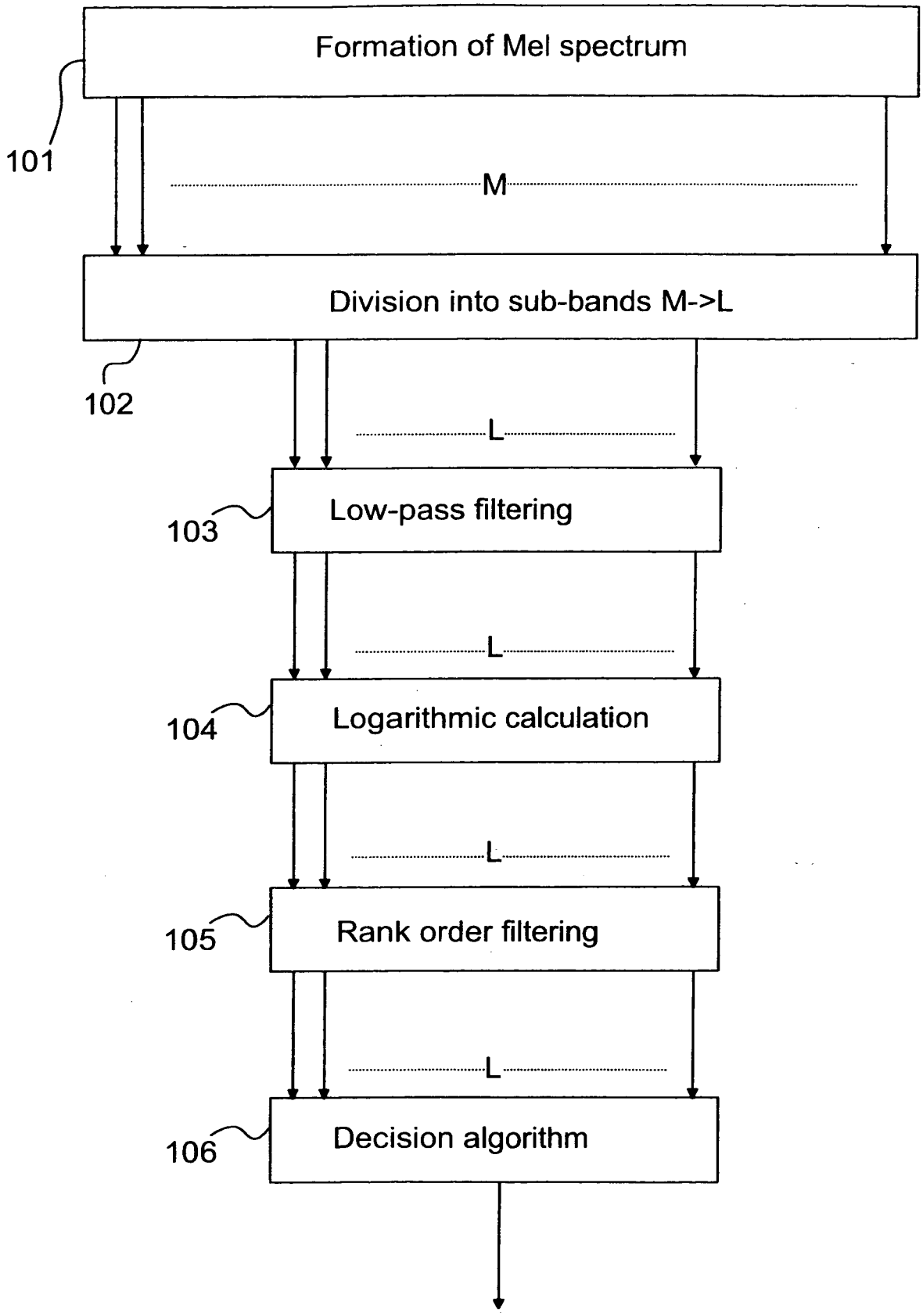

Fig 1

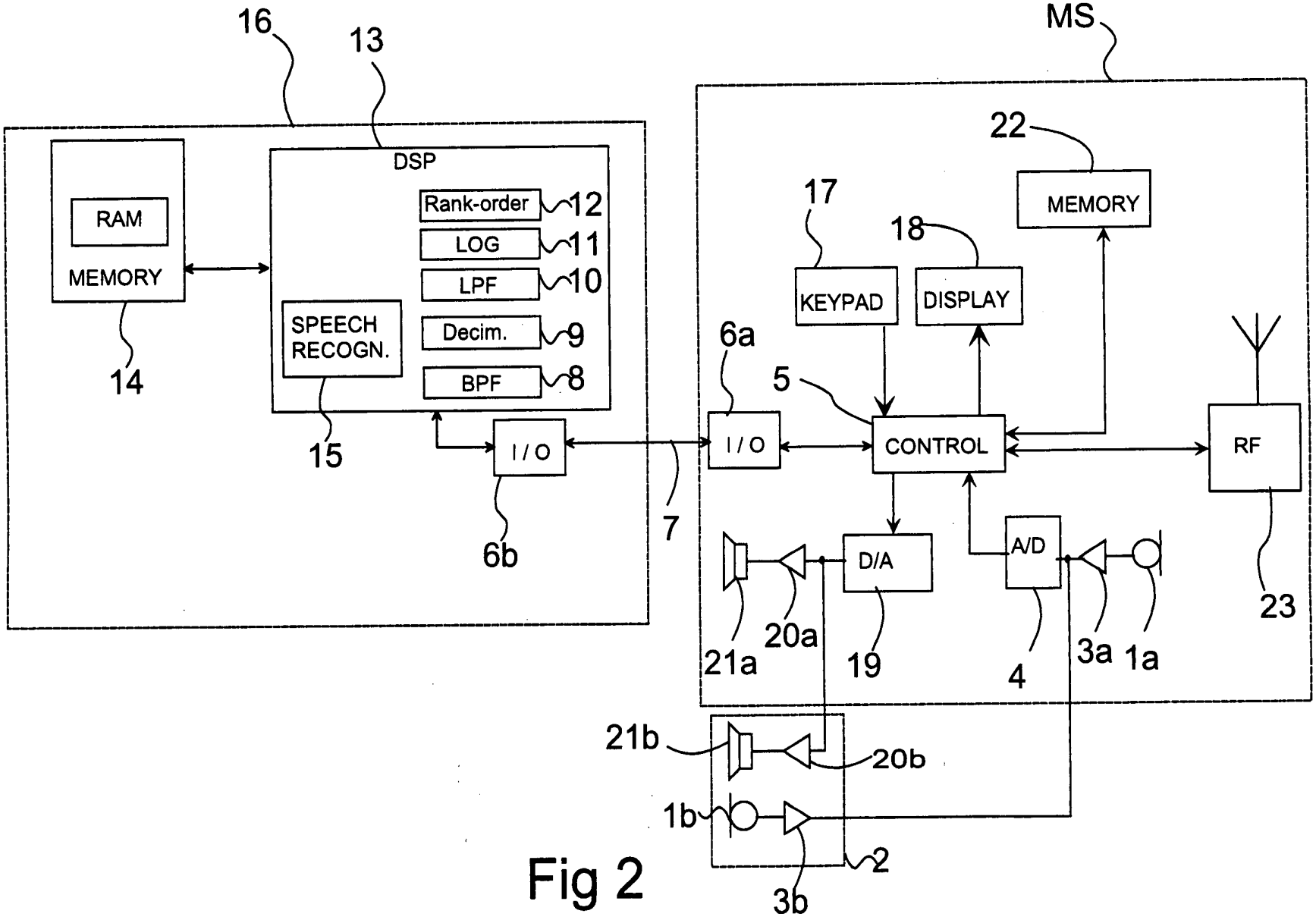
16
13
MS
22
DSP
RAM
MEMORY
14
Rank-order
12
LOG
11
LPF
10
SPEECH RECOGN.
Decim.
9
BPF
8
15
I / O
6b
7
17
18
MEMORY
KEYPAD
DISPLAY
6a
5
I / O
CONTROL
RF
D/A
A/D
21a
20a
19
4
3a
1a
23
21b
20b
1b
2
3b

Fig 2

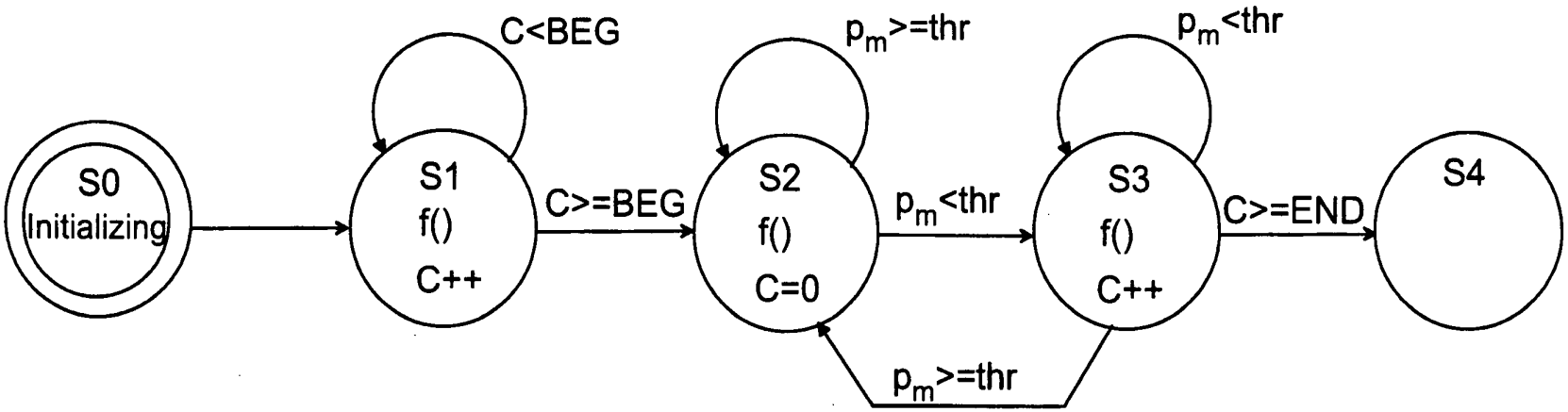
S0
Initializing
S1
f()
C++
C<BEG
C>=BEG
S2
f()
C=0
p_m>=thr
p_m<thr
S3
f()
C++
p_m<thr
C>=END
S4
p_m>=thr

Fig 3

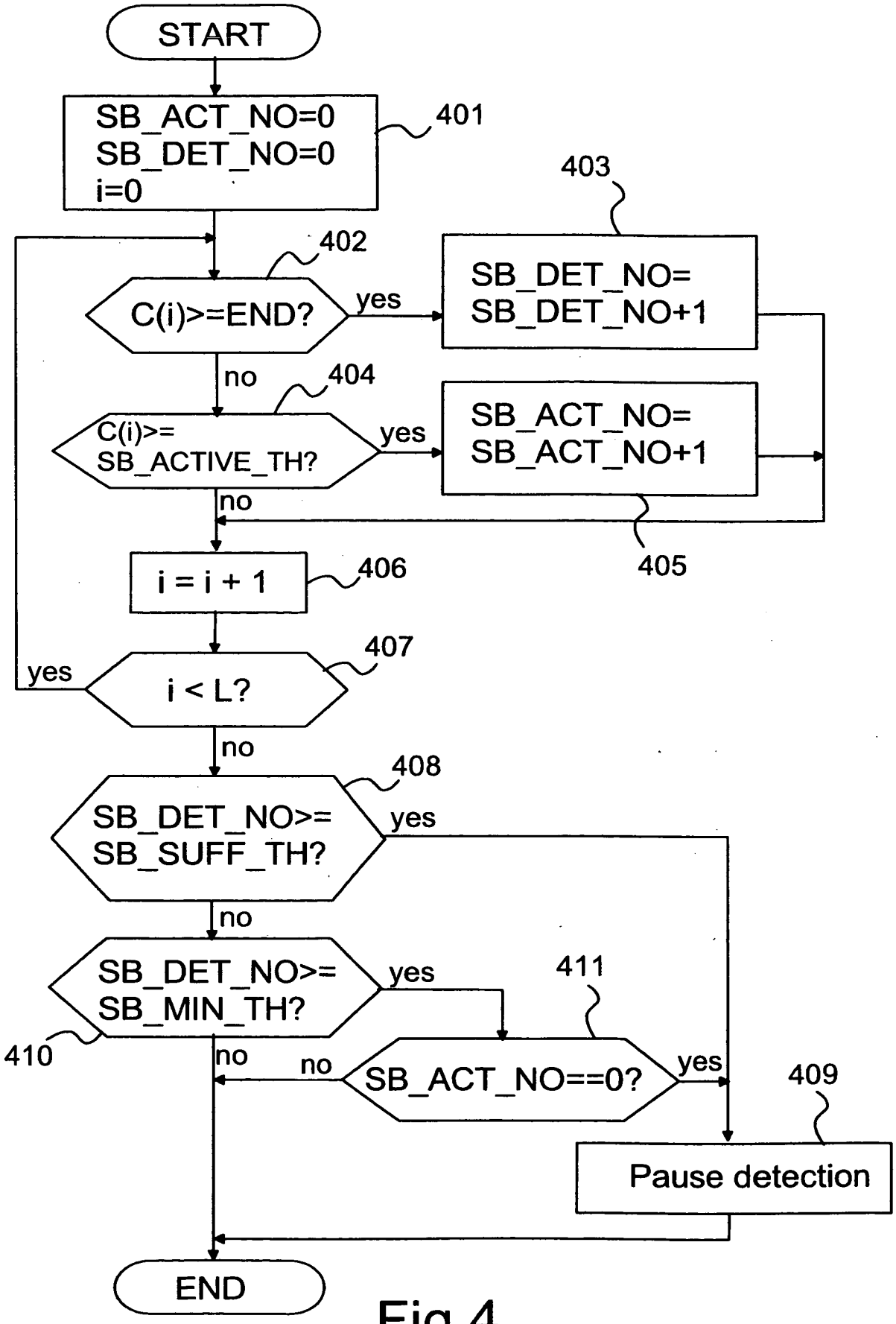
START
SB_ACT_NO=0
SB_DET_NO=0
i=0
401
403
402
C(i)>=END?
yes
SB_DET_NO=
SB_DET_NO+1
no
404
C(i)>=
SB_ACTIVE_TH?
yes
SB_ACT_NO=
SB_ACT_NO+1
no
405
i = i + 1
406
407
yes
i < L?
no
408
SB_DET_NO>=
SB_SUFF_TH?
yes
no
SB_DET_NO>=
SB_MIN_TH?
yes
411
410
no
no
SB_ACT_NO==0?
yes
409
Pause detection
END

Fig 4

PATENTTI- JA REKISTERIHALLITUS
NATIONAL BOARD OF PATENTS AND REGISTRATION

#2

Helsinki 19.11.1999

E T U O I K E U S T O D I S T U S
P R I O R I T Y D O C U M E N T

Hakija Applicant	**Nokia Mobile Phones Ltd** **Espoo**
Patenttihakemus nro Patent application no	**990078**
Tekemispäivä Filing date	**18.01.1999**
Kansainvälinen luokka International class	**G10L**

Keksinnön nimitys
Title of invention

"Menetelmä puheen tunnistuksessa ja puheentunnistuslaite"

Täten todistetaan, että oheiset asiakirjat ovat tarkkoja jäljennöksiä patentti- ja rekisterihallitukselle alkuaan annetuista selityksestä, patenttivaatimuksista, tiivistelmästä ja piirustuksista.

This is to certify that the annexed documents are true copies of the description, claims, abstract and drawings originally filed with the Finnish Patent Office.

Pirjo Kaila
Tutkimussihteeri

Maksu 300,- mk
Fee 300,- FIM

Osoite: Arkadiankatu 6 A
P.O.Box 1160
FIN-00101 Helsinki, FINLAND
Puhelin: 09 6939 500
Telephone: + 358 9 6939 500
Telefax: 09 6939 5204
Telefax: + 358 9 6939 5204

Menetelmä puheentunnistuksessa ja puheentunnistuslaite

Nyt esillä oleva keksintö kohdistuu oheisen patenttivaatimuksen 1 johdanto-osan mukaiseen menetelmään puheentunnistuksessa, oheisen patenttivaatimuksen 8 johdanto-osan mukaiseen puheentunnistuslaitteeseen ja oheisen patenttivaatimuksen 11 johdanto-osan mukaiseen puheella ohjattavaan langattomaan viestimeen.

Langattomien viestimien käytön helpottamiseksi on kehitetty puheentunnistuslaitteita, joiden avulla käyttäjä voi lausua puhekomentoja, jotka puheentunnistuslaite pyrkii tunnistamaan ja muuntamaan puhekomentoa vastaavaksi toiminnoksi, esim. puhelinnumeron valintakomennoksi. Hankaluutena puheohjauksen toteuttamisessa on mm. se, että eri käyttäjät lausuvat puhekomennot eri tavalla: puhenopeus voi olla erilainen eri käyttäjillä, samoin puheen voimakkuus, äänen sävy jne. Lisäksi puheentunnistusta häiritsee mahdoilinen taustamelu, jonka häiritsevyys ulkona ja autossa voi olla huomattavaa. Taustamelu vaikeuttaa sanojen tunnistusta sekä eri sanojen erottamista toisistaan esim. puhelinnumeroa lausuttaessa.

Joissakin puheentunnistuslaitteissa on käytetty kiinteään aika-ikkunaan perustuvaa tunnistusmenetelmää. Tällöin käyttäjällä on ennalta määrätty aika, jonka kuluessa hänen on lausuttava haluamansa komentosana. Aika-ikkunan kuluttua umpeen puheentunnistuslaite pyrkii selvittämään, minkä sanan/komennon käyttäjä lausui. Tällaiseen kiinteään aika-ikkunaan perustuvassa menetelmässä on kuitenkin mm. se epäkohta, että kaikki lausuttavat sanat eivät ole yhtä pitkiä, esim. nimien kohdalla etunimi on usein selvästi lyhyempi kuin sukunimi. Tällöin lyhyemmän sanan jälkeen kuluu enemmän aikaa tunnistukseen kuin pidemmän sanan tunnistuksessa. Tämä on epämiellyttävää käyttäjän kannalta. Lisäksi aika-ikkuna on asetettava hitaampien puhujien mukaan, ettei tunnistusta aloiteta, ennen kuin koko sana on lausuttu. Nopeammin sanoja lausuttaessa viive lausumisen ja tunnistuksen välillä lisää epämiellyttävyyden tunnetta.

Toinen tunnettu puheentunnistusmenetelmä perustuu puhesignaaleista muodostettuihin malleihin ja niiden vertailuun. Komentosanoista muo-

dostetut mallit on etukäteen tallennettu tai käyttäjä on voinut opettaa haluamiaan sanoja, joista on muodostettu ja tallennettu mallit. Puheentunnistuslaite vertailee tallennettuja malleja käyttäjän lausumista äänteistä muodostettuihin piirrevektoreihin sanojen lausumisen aikana ja laskee todennäköisyyksiä puheentunnistuslaitteen sanaston eri sanoille (komentosanoille). Todennäköisyyden ylittäessä jollakin komentosanalla ennalta asetetun arvon, puheentunnistuslaite valitsee tämän komentosanan tunnistustulokseksi. Tällöin voi virheellisiä tunnistustuloksia syntyä erityisesti sellaisten sanojen kohdalla, joissa sanan alku muistuttaa äänteellisesti jotakin muuta sanastoon kuuluvaa sanaa. Esimerkiksi käyttäjä on opettanut puheentunnistuslaitteelle sanat "Mari" ja "Marika". Jos käyttäjä lausuu sanaa "Marika", saattaa puheentunnistuslaite tehdä tunnistuspäätökseksi "Mari", vaikka käyttäjä ei olisi ehtinyt lausua vielä sanan loppua. Tällaisissa puheentunnistuslaitteissa käytetään usein ns. Hidden-Markov-Model -puheentunnistusmenetelmää (HMM).

Patentissa US-4,870,686 on esitetty puheentunnistusmenetelmä ja puheentunnistuslaite, jossa käyttäjän sanojen lopun ilmaiseminen perustuu hiljaisuuteen, siis puheentunnistuslaite tutkii, onko äänisignaalia havaittavissa vai ei. Ongelmana tässä ratkaisussa on se, että liian voimakas taustamelu voi estää taukojen havaitsemisen, jolloin puheentunnistus ei onnistu.

Nyt esillä olevan keksinnön eräänä tarkoituksena on aikaansaada parannettu menetelmä puheessa olevien taukojen havaitsemiseksi ja puheentunnistuslaite. Keksintö perustuu siihen ajatukseen, että jaetaan tutkittava äänikaista alikaistoihin ja tutkitaan signaalin tehoa kullakin alikaistalla. Mikäli riittävän usealla alikaistalla signaalin teho alittaa tietyn rajan riittävän pitkän ajan, tehdään päätelmä siitä, että puheessa on tauko. Nyt esillä olevan keksinnön mukaiselle menetelmälle on tunnusomaista se, mitä on esitetty oheisen patenttivaatimuksen 1 tunnusmerkkiosassa. Nyt esillä olevan keksinnön mukaiselle puheentunnistuslaitteelle on tunnusomaista se, mitä on esitetty oheisen patenttivaatimuksen 8 tunnusmerkkiosassa. Nyt esillä olevan keksinnön mukaiselle langattomalle viestimelle on tunnusomaista se, mitä on esitetty oheisen patenttivaatimuksen 11 tunnusmerkkiosassa.

Nyt esillä olevalla keksinnöllä saavutetaan merkittäviä etuja tunnetun tekniikan mukaisiin ratkaisuihin verrattuna. Keksinnön mukaisella menetelmällä saadaan luotettavampi sanavälin ilmaisu kuin tunnetun tekniikan mukaisilla menetelmillä. Tällöin puheentunnistuksen luotettavuus paranee ja virheellisten tunnistusten ja epäonnistuneiden tunnistusten määrä pienenee. Lisäksi puheentunnistuslaite on joustavampi erilaisten käyttäjien puhetottumusten suhteen, koska puhekomennot voidaan lausua hitaammin tai nopeammin ilman, että tunnistuksessa on epämiellyttävää viivettä tai että tunnistus tapahtuisi kesken sanan lausumisen.

Keksinnön mukaisella alikaistoihin jakamisella saadaan ulkoisten häiriöiden vaikutusta pienennettyä. Tyypillisesti häiriösignaalit esim. autossa ovat suhteellisen matalataajuisia. Tunnetun tekniikan mukaisissa ratkaisuissa koko käsiteltävän signaalin taajuusalueen sisältämää energiaa käytetään tunnistuksessa hyväksi, jolloin voimakkaat mutta kapeakaistaiset signaalit heikentävät signaali-kohinasuhdetta merkittävästi. Sen sijaan jaettaessa tutkittava taajuusalue keksinnön mukaisesti alikaistoihin, saadaan sellaisilla alikaistoilla, joilla häiritsevien signaalien osuus on suhteellisen pieni, signaali-kohinasuhdetta parannettua merkittävästi, mikä parantaa tunnistusvarmuutta.

Nyt esillä olevaa keksintöä selostetaan seuraavassa tarkemmin viitaten samalla oheisiin piirustuksiin, joissa

kuva 1 esittää vuokaaviona keksinnön erään edullisen suoritusmuodon mukaista menetelmää,

kuva 2 esittää keksinnön erään edullisen suoritusmuodon mukaista puheentunnistuslaitetta pelkistettynä lohkokaaviona,

kuva 3 esittää keksinnön erään edullisen suoritusmuodon mukaisessa menetelmässä sovellettavaa sijalukusuodatusta (rank-order filtering) tilakonekaaviona, ja

kuva 4 esittää vuokaaviona keksinnön erään edullisen suoritusmuodon mukaisessa menetelmässä sovellettavaa tauon päättelylogiikkaa.

Selostetaan seuraavassa keksinnön erään edullisen suoritusmuodon mukaisen menetelmän toimintaa viitaten samalla kuvan 1 vuokaavioon käyttäen esimerkkinä kuvan 2 lohkokaavion mukaista puheella ohjattavaa langatonta viestintä MS. Puheentunnistuksessa suoritetaan sinänsä tunnetusti akustisen signaalin (puheen) muuntaminen sähköiseksi signaaliksi mikrofonilla, kuten langattoman viestimen MS mikrofonilla 1a tai kaiutintoiminnon 2 mikrofonilla 1b. Puhesignaalin taajuusvaste rajoittuu tyypillisesti alle 10 kHz:n taajuusalueelle, esim. taajuusalueelle 100 Hz—10 kHz. Puheen taajuusvaste ei kuitenkaan ole vakio koko taajuusalueella, vaan siinä matalampia taajuuksia esiintyy enemmän kuin korkeampia taajuuksia. Lisäksi eri henkilöillä puheen taajuusvaste on erilainen. Keksinnön mukaisessa menetelmässä tutkittava taajuusalue jaetaan kapeampiin alitaajuusalueisiin (alikaistoihin, M kpl). Tätä esittää lohko 101 oheisessa kuvassa 1. Näitä alitaajuusalueita ei tehdä tasalevyisiksi, vaan puheen ominaispiirteet huomioiden, jolloin osa alitaajuusalueista on kapeampia ja osa on leveämpiä. Puheelle ominaisilla, alemmilla taajuuksilla jako on tiheämpi, eli alitaajuusalueet ovat kapeampia, kuin puheessa harvemmin esiintyvillä, korkeammilla taajuuksilla. Tähän perustuu myös sinänsä tunnettu mel-taajuusjako (Mel Frequency Scale), jossa taajuuskaistojen leveys perustuu logaritmiseen taajuuden funktioon.

Alikaistoihin jakamisen yhteydessä alikaistojen signaalit muunnetaan pienemmälle näytetaajuudelle esim. alinäytteistämällä tai alipäästösuodattamalla. Tällöin lohkosta 101 näytteitä siirretään jatkokäsittelyyn tällä alemmalla näytetaajuudella. Tämä näytetaajuus on edullisesti n. 100 Hz, mutta on selvää, että nyt esillä olevan keksinnön puitteissa myös muita näytetaajuuksia voidaan soveltaa. Näistä näytteistä muodostetaan mainittuja piirrevektoreita.

Mikrofonissa 1a, 1b muodostettu signaali vahvistetaan vahvistimessa 3a, 3b ja muunnetaan digitaaliseksi analogia-digitaalimuuntimessa 4. Analogia/digitaalimuunnoksen tarkkuus on tyypillisesti välillä 12—32 bittiä ja puhesignaalin muuntamisessa näytteitä otetaan edullisesti 8000—14000 kertaa sekunnissa, mutta keksintöä voidaan soveltaa myös muilla näytteenottonopeuksilla. Kuvan 2 langattomassa viestimessä MS näytteenotto on järjestetty suoritettavaksi kontrollerin 5 ohjaamana. Digitaalisessa muodossa oleva äänisignaali siirretään langat-

toman viestimen MS kanssa toiminnallisessa yhteydessä olevaan puheentunnistuslaitteeseen 16, jossa suoritetaan keksinnön edullisen suoritusmuodon mukaisen menetelmän eri vaiheita. Siirto suoritetaan esim. liityntälohkojen 6a, 6b ja liityntäväylän 7 kautta. Puheentunnistuslaite 16 voi käytännön sovelluksissa olla toteutettuna myös itse langattomassa viestimessä MS tai muussa puheohjattavassa laitteessa, tai erillisenä lisälaitteena tai vastaavana.

Alikaistoihin jako tehdään edullisesti ensimmäisessä suodatinlohkossa 8, johon digitaaliseksi muunnettu signaali johdetaan. Tämä ensimmäinen suodatinlohko 8 koostuu useista, tässä edullisessa suoritusmuodossa digitaalitekniikalla toteutetuista, kaistanpäästösuodattimista, joiden päästökaistan taajuusalueet sekä kaistanleveydet eroavat toisistaan. Tällöin kunkin kaistanpäästösuodattimen läpäisee alkuperäisestä signaalista kaistanpäästösuodatettu osa. Selvyyden vuoksi ei kuvassa 2 ole esitetty erillisinä näitä kaistanpäästösuodattimia. Nämä kaistanpäästösuodattimet on toteutettu edullisesti signaalinkäsittely-yksikön 13 (DSP, Digital Signal Processor) sovellusohjelmistossa, kuten on sinänsä tunnettua.

Seuraavassa vaiheessa 102 vähennetään alikaistojen lukumäärää edullisesti desimoimalla desimointilohkossa 9, jolloin muodostuu L kappaletta alikaistoja (L<M), joiden energiatasot ovat mitattavissa. Näiden alitaajuusalueiden signaalinvoimakkuuksien perusteella voidaan määrittää signaalin energia kullakin alikaistalla. Myös desimointilohko 9 voidaan toteuttaa digitaalisen signaalinkäsittely-yksikön 13 sovellusohjelmistossa.

Etu, joka saavutetaan lohkon 1 mukaisella M alikaistaan jakamisella on se, että näitä M:n eri alikaistan arvoja voidaan käyttää tunnistuksessa apuna tunnistustuloksen varmentamiseksi erityisesti sellaisessa sovelluksessa, jossa käytetään Mel-taajuusjaon mukaisia kertoimia. Lohko 101 voidaan kuitenkin toteuttaa myös siten, että siinä muodostetaan suoraan L kappaletta alikaistoja, jolloin lohkoa 102 ei tarvita.

Toisessa suodatinlohkossa 10 suoritetaan desimointivaiheessa muodostetuille alikaistojen signaaleille alipäästösuodatus (vaihe 103 kuvassa 1), jolloin lyhyet signaalinvoimakkuuden muutokset suodattuvat

ja eivät pääse vaikuttamaan merkittävästi signaalin energiatason määrittämiseen jatkossa. Suodatuksen jälkeen lasketaan lohkossa 11 kunkin alikaistan energiatasosta logaritmifunktio (vaihe 104), jonka muodostamat laskentatulokset tallennetaan jatkokäsittelyä varten muistivälineisiin 14 muodostettuihin alikaistakohtaisiin puskureihin (ei esitetty). Nämä puskurit ovat edullisesti ns. FIFO-tyyppisiä (First In - First Out), joihin laskentatulokset tallennetaan esim. 8- tai 16-bittisinä lukuina. Kuhunkin puskuriin mahtuu N kappaletta laskentatuloksia. Arvo N riippuu kulloisestakin sovelluksesta. Puskuriin tallennetut laskentatulokset p(t) kuvaavat siis alikaistan suodatettua, logaritmista energiatasoa eri mittausajanhetkinä.

Järjestelylohko 12 suorittaa laskentatuloksille ns. rank-order -suodatuksen (vaihe 105), jossa eri laskentatulosten keskinäistä suuruutta vertaillaan. Tässä vaiheessa 105 tutkitaan alikaistoittain se, onko puheessa mahdollisesti tauko. Tämä tutkiminen on esitetty tilakonekaaviona kuvassa 3. Tämän tilakoneen toiminnot toteutetaan olennaisesti samanlaisina kullekin alikaistalle. Tilakoneen eri toimintatiloja S0, S1, S2, S3 ja S4 on esitetty ympyröillä. Näiden tilaympyröiden sisään on merkitty kussakin toimintatilassa suoritettavat toimenpiteet. Nuolet 301, 302, 303, 304 ja 305 kuvaavat siirtymisiä toimintatiloista toiseen. Näiden nuolien yhteyteen on merkitty kriteerit, joiden toteutuminen aikaansaa tämän siirtymisen. Kaaret 306, 307 ja 308 kuvaavat tilannetta, jossa toimintatilaa ei vaihdeta. Myös näiden kaarien yhteyteen on merkitty kriteerit toimintatilan säilyttämiseksi ennallaan.

Toimintatiloissa S1, S2 ja S3 on esitetty funktio f(), joka tarkoittaa seuraavien toimenpiteiden suorittamista mainituissa toimintatiloissa: laskentatuloksia p(t) tallennetaan puskuriin edullisesti N kappaletta, joista etsitään pienin maksimiarvo p_min(t) ja suurin minimiarvo p_min(t) edullisesti seuraavilla kaavoilla:

$$p\_min(t) = min\Big[max\big(p(i-N+1), p(i-N+2), \ldots, p(i)\big)\Big], \quad i = N, N+1, \ldots, t$$

$$p\_max(t) = max\Big[min\big(p(i-N+1), p(i-N+2), \ldots, p(i)\big)\Big], \quad i = N, N+1, \ldots, t$$

Funktiossa f() haetaan siis maksimiarvoksi p_max(t) eri alikaistapuskureihin tallennetuista laskentatuloksista p(i) suurin minimiarvo ja mi-

nimiarvoksi p_min(t) pienin maksimiarvo. Tämän jälkeen lasketaan mediaaniteho $p(t)_m$, joka on mediaaniarvo puskuriin tallennetuista laskentatuloksista p(t) sekä kynnysarvo thr kaavalla $thr = p\_min + k \cdot (p\_max - p\_min)$, jossa 0 < k < 1. Seuraavaksi funktiossa f() suoritetaan mediaanitehon $p(t)_m$ vertailu edellä laskettuun kynnysarvoon. Vertailun tulos aikaansaa erilaisia toimenpiteitä riippuen siitä, missä toimintatilassa tilakone kulloinkin on. Tätä kuvataan jäljempänä tarkemmin eri toimintatilojen kuvauksen yhteydessä.

Sen jälkeen kun puheesta on tallennettu joukko alikaistakohtaisia laskentatuloksia p(t) (N kpl/alikaista), puheentunnistuslaite siirtyy suorittamaan mainittua tilakonetta, joka on toteutettu joko digitaalisen signaalinkäsittely-yksikön 13 tai kontrollerin 5 sovellusohjelmistossa. Ajoitus voidaan muodostaa sinänsä tunnetusti edullisesti oskillaattorilla, kuten kideoskillaattorilla (ei esitetty). Suoritus aloitetaan tilasta S0, jossa tehdään tilakoneessa käytettävien muuttujien asettamiset alkuarvoihin (init()): taukolaskuri C nollataan, tehominimiarvo p_min aloitusajanhetkellä t=1 (p_min(t=1)) asetetaan teoreettisesti arvoon ∞, käytännössä puheentunnistuslaitteessa käytettävissä olevaksi suurimmaksi mahdolliseksi lukuarvoksi. Tähän maksimiarvoon vaikuttaa se, kuinka monella bitillä näitä tehoarvoja lasketaan. Vastaavasti tehomaksimiarvo p_max aloitusajanhetkellä t=1 (p_max(t=1)) asetetaan teoreettisesti arvoon $-\infty$, käytännössä puheentunnistuslaitteessa käytettävissä olevaksi pienimmäksi mahdolliseksi lukuarvoksi.

Alkuarvojen asetuksen jälkeen toiminta siirtyy tilaan S1, jossa suoritetaan mainitun funktion f() edellä esitetyt toimenpiteet, jolloin mm. tehojen minimiarvo p_min ja maksimiarvo p_max sekä mediaaniteho $p(t)_m$ lasketaan. Toimintatilassa S1 kasvatetaan lisäksi taukolaskuria C yhdellä. Tässä toimintatilassa pysytään, kunnes ennalta määritetty alkuviive on kulunut umpeen. Tämä selvitetään vertailemalla taukolaskuria C ennalta asetettuun aloitusarvoon BEG. Siinä vaiheessa kun taukolaskuri C on saavuttanut aloitusarvon BEG, toiminta siirtyy tilaan S2.

Toimintatilassa S2 taukolaskuri C nollataan ja suoritetaan funktion f() toimenpiteet, kuten uuden laskentatuloksen p(t) tallennus, tehominimin p_min, tehomaksimin p_max ja mediaanitehon $p(t)_m$ sekä kynnysarvon thr laskenta. Laskettua kynnysarvoa ja mediaanitehoa verrataan kes-

kenään ja mikäli mediaaniteho on pienempi kuin kynnysarvo, siirrytään toimintatilaan S3, muussa tapauksessa toimintatilaa ei vaihdeta, vaan suoritetaan tämän toimintatilan S2 edellä esitetyt toimenpiteet uudelleen.

Toimintatilassa S3 kasvatetaan taukolaskuria C yhdellä ja suoritetaan funktio f(). Jos vertailu osoittaa, että mediaaniteho on edelleen pienempi kuin kynnysarvo, tutkitaan taukolaskurin C arvo sen selvittämiseksi, onko mediaaniteho ollut tietyn ajan alle tehon kynnysarvon. Tämän aikarajan täyttyminen on selvitettävissä vertaamalla taukolaskurin C arvoa ilmaisuaikarajaan END. Jos laskurin arvo on suurempi tai yhtäsuuri kuin mainittu ilmaisuaikaraja END, merkitsee se sitä, että kyseisellä alikaistalla ei puhetta ole havaittavissa, jolloin poistutaan tilakoneesta.

Jos toimintatilassa S3 kynnysarvon ja mediaanitehon vertailu kuitenkin osoitti, että mediaaniteho on ylittänyt tehon kynnysarvon, voidaan tästä tehdä päätelmä, että puhetta on tällä alikaistalla havaittavissa ja tilakone palautuu toimintatilaan S2, jossa mm. taukolaskuri C nollataan ja laskenta aloitetaan alusta.

Edellä oli siis kuvattu keksinnön erään edullisen suoritusmuodon mukaisessa menetelmässä käytettävän tilakoneen toimintaa yleisesti. Keksinnön mukaisessa puheentunnistuslaitteessa edellä esitetyt toimintavaiheet suoritetaan kunkin alikaistan osalta erikseen.

Näytteenotto puhesignaalista suoritetaan edullisesti määrävälein, jolloin vaiheet 101—104 suoritetaan kunkin piirrevektorin laskennan jälkeen, edullisesti n. 10 ms:n välein. Vastaavasti kunkin alikaistan tilakoneessa suoritetaan kulloinkin aktiivisena olevan toimintatilan mukaiset toimenpiteet kerran (yksi laskentakierros), esim. tilassa S3 kasvatetaan ao. alikanavan taukolaskuria C(s), suoritetaan funktio f(s), jossa mm. tehdään mediaanitehon ja kynnysarvon välinen vertailu ja sen perusteella joko säilytetään toimintatila ennallaan tai muutetaan toimintatilaa.

Kun kaikkien alikaistojen tilakoneiden osalta on suoritettu yksi laskentakierros, siirrytään puheentunnistuksessa vaiheeseen 106, jossa tutkitaan eri alikaistoista saadun informaation perusteella se, onko puhees-

sa havaittu riittävän pitkä tauko. Tätä vaihetta 106 on kuvattu vuokaaviona oheisessa kuvassa 4. Tutkimisen selventämiseksi määritetään muutamia vertailuarvoja, joille annetaan alkuarvot edullisesti puheentunnistuslaitteen valmistuksen yhteydessä, mutta näitä alkuarvoja voidaan tarvittaessa muuttaa kulloisenkin sovelluksen ja käyttöolosuhteiden mukaan. Näiden alkuarvojen asettamista esittää lohko 401 kuvan 4 vuokaaviossa:

- aktiivisuuskynnys SB_ACTIVE_TH, jonka arvo on suurempi kuin nolla, mutta pienempi kuin ilmaisuaikaraja END;
- ilmaisumäärä SB_SUFF_TH, jonka arvo on suurempi kuin nolla, mutta pienempi tai yhtäsuuri kuin alikaistojen lukumäärä L,
- alikaistojen minimimäärä SB_MIN_TH, jonka arvo on suurempi kuin nolla, mutta pienempi kuin ilmaisumäärä SB_SUFF_TH.

Keksinnön mukaisessa menetelmässä puheessa olevan tauon havaitsemiseksi tutkitaan, kuinka monella alikaistalla energiataso on mahdollisesti pysynyt mainitun tehon kynnysarvon alapuolella ja kuinka kauan. Kuten edellä olevasta tilakoneen toimintakuvauksesta käy ilmi, taukolaskuri C ilmaisee sen, kuinka pitkään alikaistalla on äänen energiataso ollut tehon kynnysarvon alapuolella. Tällöin tutkitaan kunkin alikaistan laskurin arvoa. Jos laskurin arvo on suurempi tai yhtä suuri kuin ilmaisuaikaraja END (lohko 402), merkitsee se sitä, että alikaistan energiataso on ollut tehon kynnysarvon alapuolella niin kauan, että päätös tauon havaitsemisesta voidaan tehdä tämän alikaistan osalta, eli muodostetaan alikanavakohtainen ilmaisu. Tällöin lohkossa 403 kasvatetaan ilmaisulaskuria SB_DET_NO edullisesti yhdellä.

Jos laskurin arvo on suurempi tai yhtä suuri kuin aktiivisuuskynnys SB_ACTIVE_TH (lohko 404), energiataso tällä alikaistalla on ollut tehon kynnysarvon thr alapuolella hetken, mutta ei vielä ilmaisuaikarajaa END vastaavaa aikaa. Tällöin lohkossa 405 kasvatetaan aktiivisuuslaskuria SB_ACT_NO edullisesti yhdellä. Muussa tapauksessa alikaistassa on joko äänisignaalia, tai äänisignaalin taso on ollut vain lyhyen ajan alle tehon kynnysarvon thr.

Seuraavaksi siirrytään lohkoon 406, jossa apumuuttujana käytettävää alikaistalaskuria i kasvatetaan yhdellä. Tämän alikaistalaskurin i arvon

Kun verta[redacted] mainittuihin taukolaskureihin [redacted] suoritettu, tutkitaan, kuinka monella alikaistalla on havaittu tauko (taukolaskuri oli suurempi tai yhtäsuuri kuin ilmaisuaikaraja END). Jos tällaisten alikaistojen lukumäärä on suurempi tai yhtäsuuri kuin ilmaisumäärä SB_SUFF_TH (lohko 408), menetelmässä päätellään, että puheessa on tauko (tauon tunnistuspäätös, lohko 409) ja voidaan siirtyä varsinaiseen puheentunnistukseen, jossa pyritään selvittämään se, mitä käyttäjä lausui. Jos sen sijaan alikaistojen lukumäärä on pienempi kuin ilmaisumäärä SB_SUFF_TH, tutkitaan, onko alikaistojen, joissa on tauko, määrä suurempi tai yhtäsuuri kuin alikaistojen minimimäärä SB_MIN_TH (lohko 410). Lohkossa 411 tutkitaan vielä, onko jokin alikaista aktiivinen (taukolaskuri oli suurempi tai yhtäsuuri kuin aktiivisuuskynnys SB_ACTIVE_TH, mutta pienempi kuin ilmaisuaikaraja END). Keksinnön mukaisessa menetelmässä tehdään tässä tilanteessa päätös siitä, että puheessa on tauko, jos mikään alikaista ei ole aktiivinen.

Kohinatilanteessa voi joillakin alikaistoilla kohina vaikuttaa siten, että ilmaisupäätöstä ei saada kaikilla alikaistoilla, vaikka puheessa olisi tauko, joka tulisi ilmaista. Tällöin mainitun alikaistojen minimimäärän SB_MIN_TH avulla voidaan puheessa olevan tauon ilmaisua varmentaa erityisesti kohinaisissa olosuhteissa. Tällöin kohinatilanteessa, mikäli tauko havaitaan vähintään mainitulla minimimäärällä SB_MIN_TH alikaistoja, todetaan puheessa oleva tauko, jos tauon havaitsemispäätös näillä alikaistoilla pysyy voimassa mainitun ilmaisuaikarajan END verran.

Vastaavasti hyvissä olosuhteissa mainitun ilmaisuaikarajan END käyttämisellä voidaan estää liian nopea tauon ilmaisupäätös. Hyvissä olosuhteissa voi mainitulla minimimäärällä alikaistoja tauon ilmaisupäätös tulla hyvinkin nopeasti, vaikka puheessa ei olisi sellaista taukoa, joka tulisi ilmaista. Odottamalla olennaisesti kaikkien alikanavien osalta ilmaisuaikarajan verran varmennetaan sitä, että puheessa todella on tauko.

Keksinnön eräässä toisessa edullisessa suoritusmuodossa ei ennen tauon tunnistuspäätöksen tekemistä tutkita sitä, onko jokin alikaista aktiivinen. Tällöin tauon tunnistuspäätös tehdään edellä esitettyjen vertailujen tuloksien perusteella.

Edellä esitetyt toiminnot voidaan edullisesti toteuttaa esimerkiksi puheentunnistuslaitteen kontrollerin tai digitaalisen signaalinkäsittely-yksikön sovellusohjelmistossa.

Edellä esitettyä keksinnön edullisen suoritusmuodon mukaista menetelmää puheessa olevan tauon ilmaisemiseksi voidaan soveltaa puheentunnistuslaitteen opetusvaiheessa sekä puheentunnistusvaiheessa. Opetusvaiheessa voidaan häiriöolosuhteet pitää tavallisesti suhteellisen vakioina. Sen sijaan käytettäessä puheella ohjattavaa laitetta voi taustamelun ja muiden häiriöiden määrä vaihdella huomattavasti. Puheentunnistuksen luotettavuuden parantamiseksi erityisesti vaihtelevissa olosuhteissa on keksinnön erään toisen edullisen suoritusmuodon mukaiseen menetelmään lisätty adaptiivisuutta kynnysarvon thr laskentaan. Tämän adaptiivisuuden aikaansaamiseksi käytetään muutoskerrointa UPDATE_C, jonka arvo on edullisesti suurempi kuin nolla ja pienempi kuin yksi. Muutoskertoimelle määritetään aluksi jokin alkuarvo mainitulta arvoalueelta. Tätä muutoskerrointa päivitetään puheentunnistuksen aikana edullisesti seuraavasti. Alikaistoista puskureihin tallennettujen näytteiden perusteella lasketaan suurin tehotaso win_max ja pienin tehotaso win_min. Tämän jälkeen suoritetaan mainitun lasketun suurimman tehotason win_max vertailu sen hetkiseen tehomaksimiin p_max ja mainitun lasketun pienimmän tehotason win_min vertailu tehominimiin p_min. Jos lasketun suurimman tehotason win_max ja tehomaksimin p_max välisen eron itseisarvo tai tehominimin p_min ja mainitun lasketun pienimmän tehotason win_min välisen eron itseisarvo on kasvanut edellisestä laskentakerrasta, kasvatetaan muutoskerrointa UPDATE_C. Vastaavasti jos lasketun suurimman tehotason win_max ja tehomaksimin p_max välisen eron itseisarvo tai tehominimin p_min ja mainitun lasketun pienimmän tehotason win_min välisen eron itseisarvo on pienentynyt edellisestä laskentakerrasta, pienennetään muutoskerrointa UPDATE_C. Tämän jälkeen lasketaan uusi tehomaksimi ja tehominimi seuraavasti:

$$p\_min(t) = (1 - UPDATE\_C) \cdot p\_min(t-1) + (UPDATE\_C \cdot win\_min)$$
$$p\_max(t) = (1 - UPDATE\_C) \cdot p\_max(t-1) + (UPDATE\_C \cdot win\_max)$$

Laskettuja uusia tehomaksimi- ja tehominimiarvoja käytetään seuraavalla näytteenottokierroksella mm. funktion f() suorituksen yhteydessä. Tämän adaptiivisen kertoimen määrityksen etuna on mm. se, että ympäristöolosuhteissa tapahtuvat muutokset voidaan paremmin huomioida puheentunnistuksessa ja tauon ilmaisu saadaan luotettavammaksi.

Edellä esitetyt eri toiminnot puheessa olevan tauon ilmaisemiseksi voidaan suurelta osin toteuttaa puheentunnistuslaitteen kontrollerin ja/tai digitaalisen signaalinkäsittelylaitteen sovellusohjelmistossa. Keksinnön mukaisessa puheentunnistuslaitteessa voidaan osa toiminnoista, kuten alikaistoihin jako toteuttaa myös analogiatekniikalla, kuten on sinänsä tunnettua. Menetelmän suorituksen yhteydessä voidaan eri vaiheissa muodostettavien laskentatulosten, muuttujien jne. tallennuksessa käyttää puheentunnistuslaitteen muistivälineitä 14, edullisesti luku/kirjoitusmuistia (RAM, Random Access Memory), haihtumatonta, uudelleen kirjoitettavissa olevaa lukumuistia (NVRAM, Non-Volatile RAM), FLASH-muistia jne. Myös langattoman viestimen muistivälineitä 22 voidaan käyttää tietojen tallennuksessa.

Kuvassa 2 keksinnön edullisen suoritusmuodon mukaisesta langattomasta viestimestä MS on esitetty vielä sinänsä tunnetut näppäimistö 17, näyttölaite 18, digitaali/analogiamuunnin 19, kuulokevahvistin 20a, kuuloke 21a, kaiutintoiminnon 2 kuulokevahvistin 20b, kuuloke 21b sekä suurtaajuuslohko 23.

Nyt esillä olevaa keksintöä voidaan soveltaa useiden eri periaatteella toimivien puheentunnistusjärjestelmien yhteydessä. Keksintö parantaa puheessa olevien taukokohtien ilmaisuvarmuutta, mikä varmentaa varsinaisen puheentunnistuksen tunnistusvarmuutta. Keksinnön mukaista menetelmää käytettäessä ei puheentunnistusta ole tarve suorittaa kiinteään aikaikkunaan sidottuna, joten tunnistusviive ei olennaisesti riipu siitä, kuinka nopeasti käyttäjä lausuu puhekomentoja. Myös taustamelun vaikutus puheentunnistukseen saadaan keksinnön mukaista menetelmää sovellettaessa pienemmäksi kuin tunnetun tekniikan mukaisissa puheentunnistuslaitteissa on mahdollista.

On selvää, että keksintöä ei ole rajoitettu ainoastaan edellä esitettyihin suoritusmuotoihin, vaan sitä voidaan muunnella oheisten patenttivaatimusten puitteissa.

Patenttivaatimukset:

1. Menetelmä puheentunnistuksessa puheessa olevien taukojen ilmaisemiseksi, jossa menetelmässä käyttäjän lausumien puhekomentojen tunnistamiseksi ääni muunnetaan sähköiseksi signaaliksi, **tunnettu** siitä, että menetelmässä sähköisen signaalin taajuusspektri jaetaan kahdeksi tai useammaksi alikaistaksi, tallennetaan alikaistojen signaaleista näytteitä väliajoin, määritetään alikaistojen energiatasot tallennettujen näytteiden perusteella, määritetään tehon kynnysarvo (thr), ja verrataan alikaistojen energiatasoja mainittuun tehon kynnysarvoon (thr), jolloin vertailutuloksia käytetään tauon ilmaisutuloksen muodostuksessa.

2. Patenttivaatimuksen 1 mukainen menetelmä, **tunnettu** siitä, että määritetään ilmaisuaikaraja (END) ja ilmaisumäärä (SB_SUFF_TH), jolloin menetelmässä alikanavan tauon pituuden laskenta aloitetaan alikaistan energiatason alittaessa mainitun tehon kynnysarvon (thr), jolloin menetelmässä muodostetaan alikanavakohtainen ilmaisu laskennan saavuttaessa ilmaisuaikarajan (END), tutkitaan, kuinka monella alikaistalla energiataso on ollut tehon kynnysarvon (thr) alapuolella pidempään kuin ilmaisuaikaraja (END), jolloin tauon ilmaisupäätös tehdään, jos alikanavakohtaisten ilmaisujen lukumäärä on suurempi tai yhtä suuri kuin ilmaisumäärä (SB_SUFF_TH).

3. Patenttivaatimuksen 2 mukainen menetelmä, **tunnettu** siitä, että menetelmässä lisäksi määritetään aktiivisuusaikaraja (SB_ACTIVE_TH) ja aktiivisuusmäärä (SB_MIN_TH), jolloin tauon ilmaisupäätös tehdään, jos alikanavakohtaisten ilmaisujen lukumäärä on suurempi tai yhtäsuuri kuin aktiivisuusmäärä (SB_MIN_TH), ja muilla alikanavilla alikanavan tauon pituuden laskennassa ei ole saavutettu aktiivisuusaikarajaa (SB_ACTIVE_TH).

4. Patenttivaatimuksen 1, 2 tai 3 mukainen menetelmä, **tunnettu** siitä, että tehon kynnysarvo (thr) lasketaan kaavalla

$$thr = p\_min + k \cdot (p\_max - p\_min),$$ jossa

p_min = alikanavien tallennetuista näytteistä määritetty pienin tehomaksimi, ja

p_max = alikanavien tallennetuista näytteistä määritetty suurin tehominimi.

5. Jonkin patenttivaatimuksen 1—4 mukainen menetelmä, **tunnettu** siitä, että mainittu tehon kynnysarvo (thr) lasketaan adaptiivisesti huomioimalla kulloinenkin ympäristön häiriöäänitaso.

6. Patenttivaatimuksen 5 mukainen menetelmä, **tunnettu** siitä, että mainitun tehon kynnysarvon (thr) laskemiseksi väliajoin (t) määritetään muutoskerroin (UPDATE_C), ja tallennettujen näytteiden perusteella lasketaan alikaistojen suurin tehotaso (win_max) ja pienin tehotaso (win_min), jolloin määritetään tehomaksimi (p_max) ja tehominimi (p_min) kaavoilla:

$$p\_max(i,t) = (1 - UPDATE\_C) \cdot p\_max(i,t-1) + (UPDATE\_C \cdot win\_max)$$
$$p\_min(i,t) = (1 - UPDATE\_C) \cdot p\_min(i,t-1) + (UPDATE\_C \cdot win\_min)$$

jossa $0 < UPDATE\_C < 1$,
$0 < i < L$, ja
L on alikaistojen lukumäärä

7. Patenttivaatimuksen 6 mukainen menetelmä, **tunnettu** siitä, että menetelmässä lisäksi:

- kasvatetaan muutoskerrointa (UPDATE_C), mikäli mainitun lasketun suurimman tehotason (win_max) ja tehomaksimin (p_max) välisen eron itseisarvo tai tehominimin (p_min) ja mainitun lasketun pienimmän tehotason (win_min) välisen eron itseisarvo on kasvanut,
- pienennetään muutoskerrointa (UPDATE_C), mikäli mainitun lasketun suurimman tehotason (win_max) ja tehomaksimin (p_max) välisen eron itseisarvo tai tehominimin (p_min) ja mainitun lasketun pienimmän tehotason(win_min) välisen eron itseisarvo on pienentynyt.

8. Puheentunnistuslaite (16), joka käsittää välineet (1a, 1b) käyttäjän lausumien puhekomentojen muuntamiseksi sähköiseksi signaaliksi, **tunnettu** siitä, että se käsittää lisäksi:

- välineet (8) sähköisen signaalin taajuusspektrin jakamiseksi kahdeksi tai useammaksi alikaistaksi,
- välineet (14) näytteiden tallentamiseksi väliajoin alikaistojen signaaleista,
- välineet (5, 13) energiatasojen määrittämiseksi alikaistoista tallennettujen näytteiden perusteella,
- välineet (5, 13) tehon kynnysarvon (thr) määrittämiseksi,
- välineet (5, 13) alikaistojen energiatasojen vertailemiseksi mainittuun tehon kynnysarvoon (thr), ja
- välineet (5, 13) puheessa olevan tauon ilmaisemiseksi mainittujen vertailutulosten perusteella.

9. Patenttivaatimuksen 8 mukainen puheentunnistuslaite (16), **tunnettu** siitä, että tehon kynnysarvo (thr) on laskettu kaavalla

$thr = p\_min + k \cdot (p\_max - p\_min)$, jossa

p_min = alikanavien tallennetuista näytteistä määritetty pienin tehomaksimi, ja

p_max = alikanavien tallennetuista näytteistä määritetty suurin tehominimi.

10. Patenttivaatimuksen 8 tai 9 mukainen puheentunnistuslaite (16), **tunnettu** siitä, että se käsittää lisäksi välineet (10, 11) alikaistojen signaalien suodattamiseksi ennen tallennusta.

11. Langaton viestin (MS), joka käsittää välineet (16) puheen tunnistamiseksi, ja välineet (1a, 1b) käyttäjän lausumien puhekomentojen muuntamiseksi sähköiseksi signaaliksi, **tunnettu** siitä, että välineet (16) puheen tunnistamiseksi käsittää lisäksi:

- välineet (8) sähköisen signaalin taajuusspektrin jakamiseksi kahdeksi tai useammaksi alikaistaksi,
- välineet (14) näytteiden tallentamiseksi väliajoin alikaistojen signaaleista,

- välineet (5, 13) energiatasojen määrittämiseksi alikaistoista tallennettujen näytteiden perusteella,
- välineet (5, 13) tehon kynnysarvon (thr) määrittämiseksi,
- välineet (5, 13) alikaistojen energiatasojen vertailemiseksi mainittuun tehon kynnysarvoon (thr), ja
- välineet (5, 13) puheessa olevan tauon ilmaisemiseksi mainittujen vertailutulosten perusteella.

18

13

(57) Tiivistelmä

Menetelmässä puheessa olevien taukojen ilmaisemiseksi käyttäjän lausumien puhekomentojen tunnistamista varten ääni muunnetaan sähköiseksi signaaliksi, jonka taajuusspektri jaetaan kahdeksi tai useammaksi alikaistaksi. Alikaistojen signaaleista tallennetaan näytteitä väliajoin, määritetään alikaistojen energiatasot tallennettujen näytteiden perusteella, määritetään tehon kynnysarvo (thr), ja verrataan alikaistojen energiatasoja mainittuun tehon kynnysarvoon (thr). Vertailutuloksia käytetään tauon ilmaisutuloksen muodostuksessa.

Fig. 1

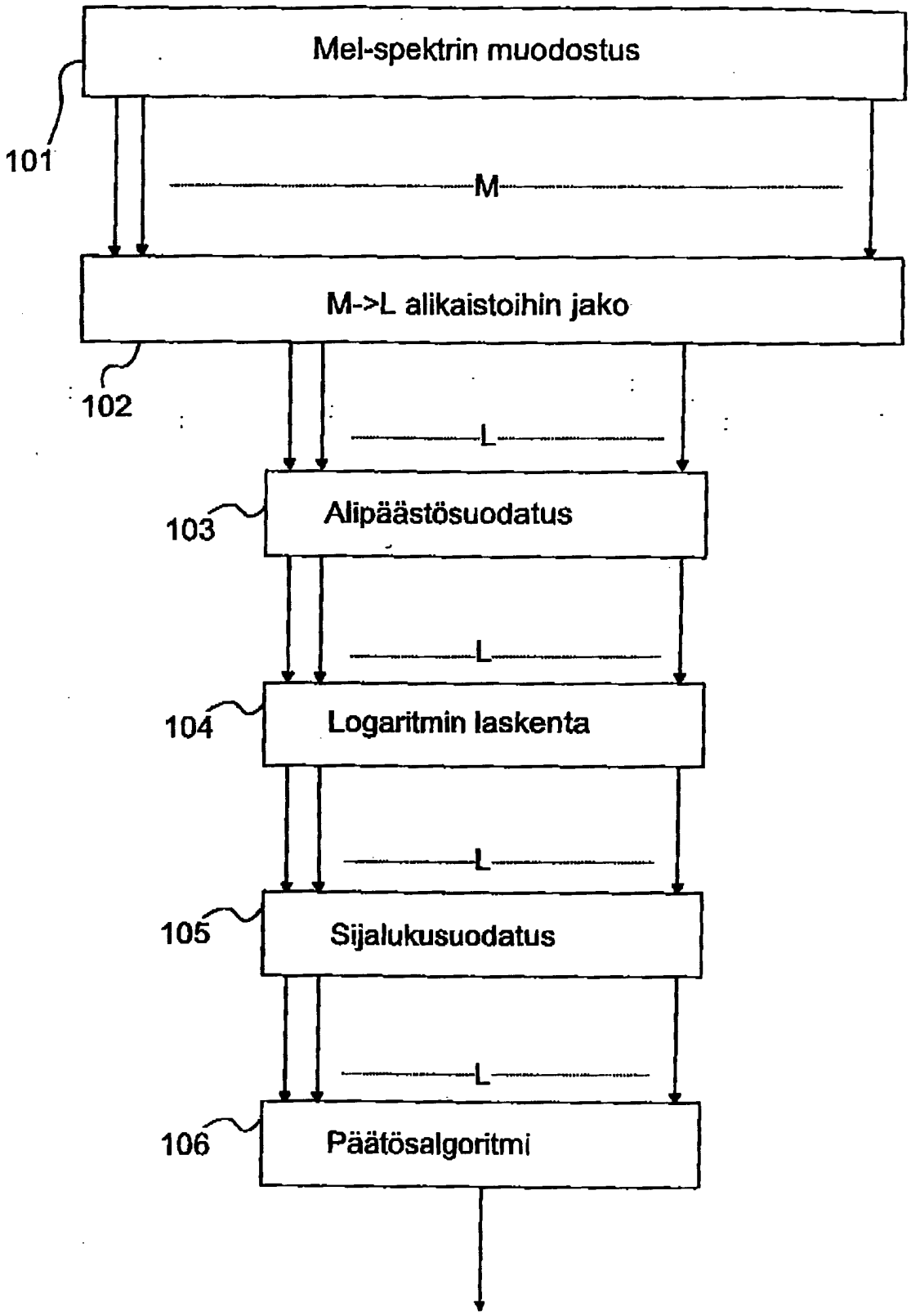
Mel-spektrin muodostus
101
M
M->L alikaistoihin jako
102
L
103
Alipäästösuodatus
L
104
Logaritmin laskenta
L
105
Sijalukusuodatus
L
106
Päätösalgoritmi

Fig 1

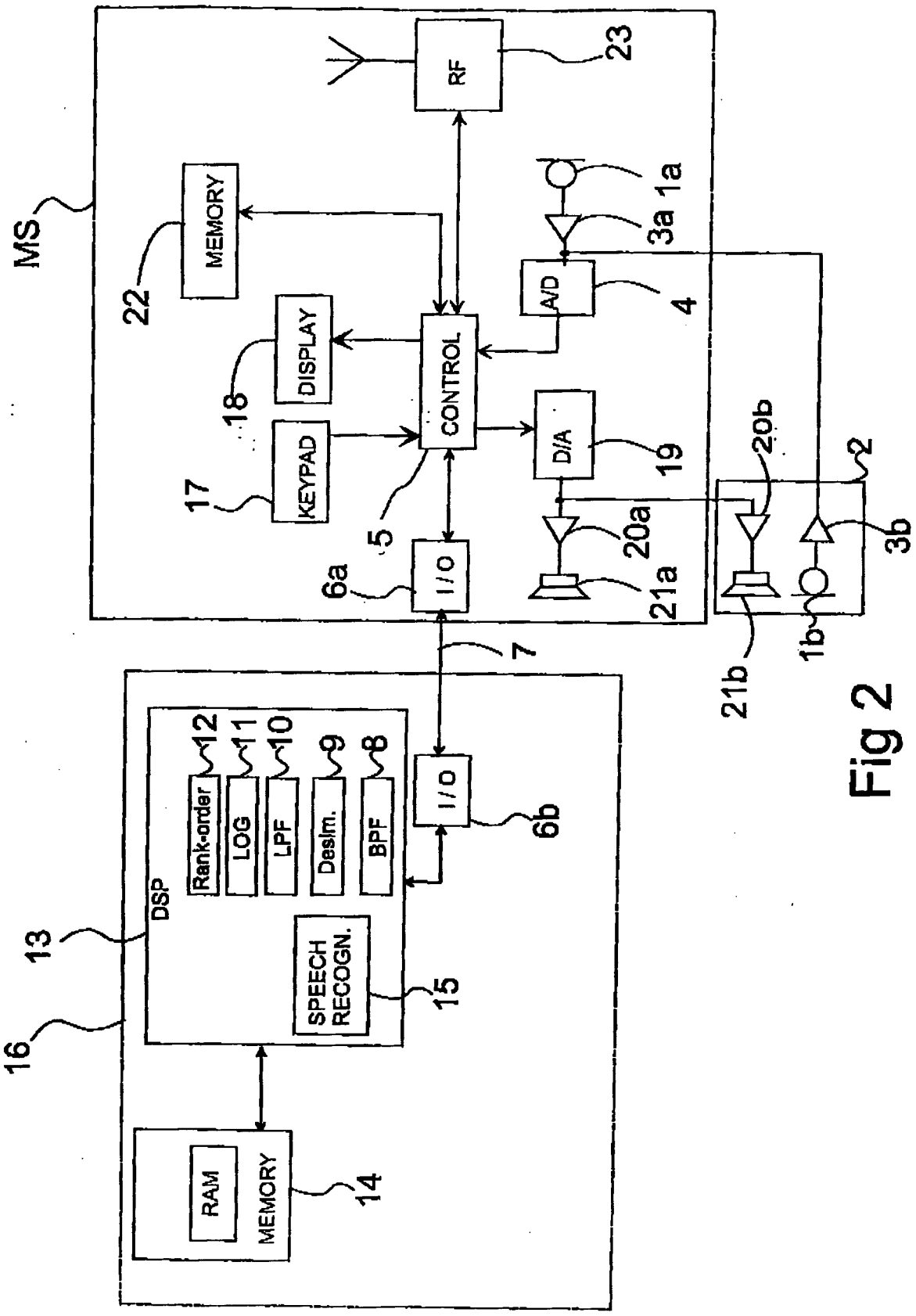
MS
RF
23
MEMORY
22
1a
3a
A/D
4
DISPLAY
18
CONTROL
KEYPAD
17
D/A
19
20b
2
5
20a
3b
6a
I/O
21a
7
21b
1b
12
11
10
9
8
Rank-order
LOG
LPF
Deslm.
BPF
I/O
6b
DSP
13
SPEECH RECOGN.
15
16
RAM
MEMORY
14

Fig 2

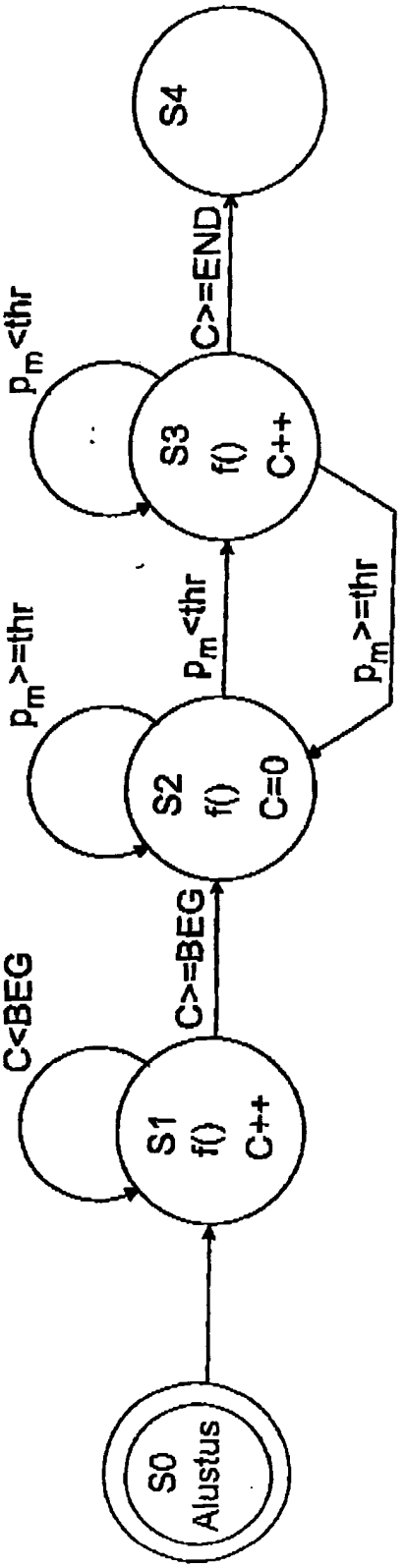
S0
Alustus
S1
f()
C++
S2
f()
C=0
S3
f()
C++
S4
C<BEG
C>=BEG
C>=END
p_m>=thr
p_m<thr
p_m>=thr
p_m<thr

Fig 3

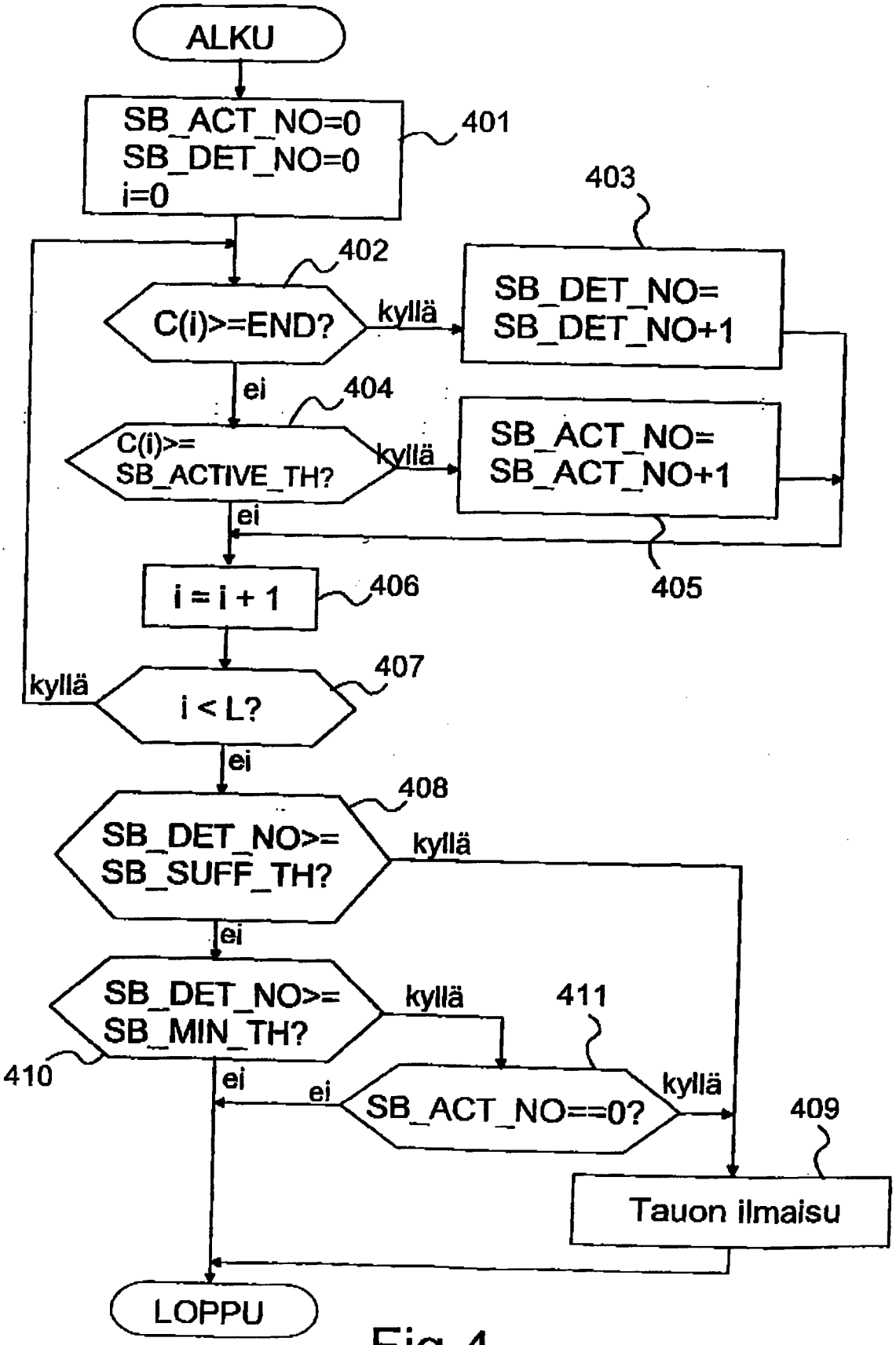
ALKU
SB_ACT_NO=0
SB_DET_NO=0
i=0
401
403
402
C(i)>=END?
kyllä
SB_DET_NO=
SB_DET_NO+1
ei
404
C(i)>=
SB_ACTIVE_TH?
kyllä
SB_ACT_NO=
SB_ACT_NO+1
ei
405
i = i + 1
406
kyllä
407
i < L?
ei
408
SB_DET_NO>=
SB_SUFF_TH?
kyllä
ei
SB_DET_NO>=
SB_MIN_TH?
kyllä
411
410
ei
ei
SB_ACT_NO==0?
kyllä
409
Tauon ilmaisu
LOPPU

Fig 4